



Data Sovereignty and Open Sharing: Reconceiving Benefit-Sharing and Governance of Digital Sequence Information

ESSAY

MASANORI ARITA 

 ubiquity press

ABSTRACT

There are ethical, legal, and governance challenges surrounding data, particularly in the context of digital sequence information (DSI) on genetic resources. I focus on the shift in the international framework, as exemplified by the CBD-COP15 decision on benefit-sharing from DSI and discuss the growing significance of data sovereignty in the age of AI and synthetic biology. Using the example of the COVID-19 pandemic, the tension between open science principles and data control rights is explained. This opinion also highlights the importance of inclusive and equitable data sharing frameworks that respect both privacy and sovereign data rights, stressing the need for international cooperation and equitable access to data to reduce global inequalities in scientific and technological advancement.

CORRESPONDING AUTHOR:

Masanori Arita

National Institute of Genetics,
RIKEN Center for Sustainable
Resource Science, JP

arita@nig.ac.jp

KEYWORDS:

Convention on Biological
Diversity (CBD); Digital
Sequence Information (DSI);
COVID-19; Open Science

TO CITE THIS ARTICLE:

Arita, M 2025 Data
Sovereignty and Open Sharing:
Reconceiving Benefit-Sharing
and Governance of Digital
Sequence Information. *Data
Science Journal*, 24: 9, pp. 1–5.
DOI: [https://doi.org/10.5334/
dsj-2025-009](https://doi.org/10.5334/dsj-2025-009)

In 2006, Clive Humby, a British mathematician in data science, coined the phrase ‘data is the new oil’ (Arthur, 2023). Like oil, data is useless in its raw state. It must be classified and analyzed to transform itself into valuables. Being an intangible object, data is not subject to any legal ownership, possession, or usufruct. Its practical usefulness comes from the process of refinement and interpretation. In other words, the value of big data is found in its potential. However, because of high expectations and potential risks, many people emphasize the importance of data control rights. One example is the data subject rights listed in the General Data Protection Regulation in Europe (GDPR, 2016). Another example is the licenses granted to computer programs or publicly available data on the Internet.

In discussing data control rights, an important distinction is between personal and non-personal data. Personal data is legally protected in many ways in each country such as in the GDPR in Europe and the Health Insurance Portability and Accountability Act (HIPAA) in the United States (Alder, 2024). Japan also has a revised Personal Data Protection Law, which defines personal information and personal data (Act 57-2003). Since the policies of these privacy laws vary widely from country to country, their generalization is beyond the scope of this contribution. For this reason, the discussion here will focus on data that does not relate to individuals.

To protect data other than personal information, patent rights and copyrights have been installed traditionally. Data is usually not copyrighted, but in some countries, databases and computer programs can become copyrighted or patented, as was discussed in World Intellectual Property Organization (WIPO website). In general, it has been difficult to imagine that such rights can be claimed for nature-derived data which does not involve human intervention, such as genome sequences. However, a major turning point has occurred at COP15, the 2022 meeting of the Conference of the Parties (COP) to the Convention on Biological Diversity (CBD Website). This conference concluded that, although not legally binding, the data control rights on genetic resources belong to the sovereign states.

BACKGROUND OF THE COP15 DECISION

The Convention on Biological Diversity (CBD) is an international treaty with the three main goals:

- (1) The conservation of biological diversity
- (2) The sustainable use of the components of biological diversity
- (3) The fair and equitable sharing of the benefits arising out of the utilization of genetic resources

With the exception of the United States, almost all countries have ratified the treaty. The parties to the treaty organize meetings every two years called the COP. The Nagoya Protocol agreed to, at COP10 in 2010, detailed procedures for implementing the ‘fair and equitable sharing of benefits’ mentioned in the third paragraph of the CBD agreement. This legally binding protocol implements the procedures exchanged between donor and user countries of genetic materials to realize the access and benefit-sharing (ABS), specifically the procedures for concluding prior informed consent (PIC) and mutually agreed terms (MAT).

COP15, which marks 10 years since the COP10 in Nagoya, was originally scheduled to be held in Kunming, China in 2020. However, due to the outbreak of the SARS-CoV-2 in early 2020, the meeting was postponed for two years, rescheduled for the end of 2022 in Montreal, where the CBD Secretariat is located. A number of landmark agreements were reached there; the ‘30 by 30’ mission of protecting at least 30% of the land and oceans by 2030, and the ‘nature positive’ mission of moving from loss to increase in biodiversity are exemplary. What was not widely reported at the time of COP15, but agreed upon at the same time, was the sharing of benefits not only from genetic materials but also from their digital sequence information (DSI), including genome sequences. It is no coincidence that remarkable progress was made at COP15. In the course of the SARS-CoV-2 pandemic, the disparity between developed and developing countries became markedly visible. The universal desire to eliminate such disparities may have accelerated the benefit-sharing from nature-derived digital information.

COVID-19 ACCELERATED THE BENEFIT SHARING FROM DATA

There is no clear consensus on what DSI indicates, even among the delegates of the COP. Its definition is deferred as the topic of future meetings. Still at least, it includes DNA and RNA sequences from natural bioresources. With this in mind, prime example of profit from DSI is the mRNA vaccine, commercialized in the wake of the SARS-CoV-2 pandemic. This vaccine can be produced solely from the genomic information of the virus and does not require the real virus particles. In the past, it has been argued that synthetic biology could make large profits from DSI alone. However, until now, there have been few examples of its realization as a huge industry. Then mRNA vaccines prevailed, as hallmark products from genome information alone.

The GISAID was the data repository recommended by the World Health Organization (WHO) to deposit SARS-CoV-2 genomes in the wake of the pandemic. This repository does not allow free public access to its data, in contrast to the conventional life science research. Instead, its data access is restricted, including prohibiting data sharing with third parties. Only under these strict conditions have many developing countries begun to participate in global genome surveillance. In fact, the early detection of Omicron and other alarming variants became possible with the worldwide coverage through GISAID. In other words, it demonstrated that global surveillance could function if the system guarantees 'benefit sharing,' including credits to data providers. On a global perspective, most countries do not wish free distribution of data. Rather, they want to ensure control over their own contribution (BEIS, 2022; Sheehan, 2024).

Although global genomic surveillance was achieved, the existing framework for vaccine distribution did not function adequately throughout the pandemic. Moreover, economic losses due to disinformation and misinformation were frequent (BEIS, 2022). It is because of those reflections that the COP15 in 2022 concluded to implement benefit sharing from DSI of genetic resources as well.

To implement benefit sharing from DSI, it is necessary to understand how DSI is used. Officially, however, there is no clear definition of DSI. Stakeholders are therefore required to prepare all patterns from its minimum interpretation as DNA sequences to a larger interpretation of genes and their products. To understand DSI usage, it seems necessary to monitor everything from data distribution channels to use cases, but foreseeing science is impossible. In the first place, what is the definition of 'benefits' arising from the utilization of data? Could monetary benefits include discounted fees? Could non-monetary benefits include public outreach and education? These discussions were postponed until COP16 in Cali, Columbia, and the discussion is still ongoing. During the two years following the COP15 decision, the CBD Secretariat organized an Informal Advisory Group led by academia to gather a wide range of opinions and attempted to delineate problems. The author has participated in these discussions as a member of the INSDC (Karsch-Mizrachi, 2024) and has argued that unrestricted data access is essential for science and education. Eventually, it was decided at COP16 that public databases and academic institutions would not be required to share monetary benefits. However, benefit sharing from the commercial sector is inevitable, and the criteria and methods of payment will be discussed in the future.

COEXISTENCE OF OPEN SCIENCE AND DATA CONTROL RIGHTS

In the movement of open science, the discussion on data control has been centered around privacy. In addition to this, we now need to consider the rights of sovereign states for nature-derived DSI. The fundamental reason for this shift is that the traditional understanding of open data and sharing may cause friction not only with the principles of privacy, but also with the principles of inclusion and diversity. To achieve equitable benefit sharing from data, advanced infrastructure, technology, and knowledge need to be all shared. Because economies of scaling are strongly at play, however, even if every player releases data equally for free, the benefits of open data will not be distributed to the providers in proportion to the amount they contribute; the player that can process data more efficiently will likely become the sole winner. This is why many developing countries cooperated with the GISAID repository, which did not allow open data during the SARS-CoV-2 pandemic, opposing the framework of unrestricted data access that developed countries advocated.

One of the key principles of the sustainable development goals is 'leave no one behind.' To achieve this, equitable utilization of data must be realized. But what kind of data usage can

be considered equitable? The current management policy may be fair but not equitable. The UNESCO philosophy of open science therefore includes the realization of openness not only in data and knowledge but also in hardware and infrastructure while maintaining inclusion and diversity. The measures to realize it will differ from time to time and from country to country. New indices such as Knowledge Sharing Index and Capacity Building Index by UNESCO help to understand the diversity in the world ([UNESCO website](#)). Aiming for equitable open science will continue to be an important issue to be solved globally.

CROSSROAD OF PERSONAL AND NON-PERSONAL DATA

The rapid advances of AI can blur and complicate the current distinction between personal and non-personal data. For example, a more powerful technology may reveal personal details from the combination of anonymized or non-personal data. For example, an anonymized dataset from the New York City Taxi and Limousine Commission was associated with other internet contents to reveal specific taxi rides by celebrities ([Atokar, 2014](#)). Likewise, creation of synthetic bio-species may raise not only ethical questions but also economic ones when they produce sizable consequences such as environmental changes. Thus, technological advancements can rapidly alter the landscape of data utilization. At the same time, we need to be aware that the notion of 'personal' is strongly influenced by Western culture. In many indigenous and local communities, traditional knowledge has been locally or communally shared, and this sharing style is not considered 'personal' as understood in Western context. In the spirit of inclusion, we need to consider more about adding diverse perspectives in the decision-making and data governance framework.

CONCLUSIONS

Data ethics has become increasingly important with the advancement of AI. While many ethical debates center around privacy, we now need to consider environmental ethics in connection with benefit-sharing across countries. For example, it is important to assess the habitats of endangered biological species accurately, but inadvertent disclosure of their geographical location may induce overexploitation. Many biological species are identified through their genome analysis, but publication of all data without restriction allows free utilization of their DSI and may exacerbate exploitation of DSI and traditional knowledge. Likewise, open data for pathogens may assist bioterrorism. Today, basic science is tightly knitted with national economy and political balance, and basic biology is no exception. We need to be aware of the national policies of related countries when we deal with DSI of bioresources. International multilateral mechanisms such as WIPO and WHO will become more important to lessen global inequalities in data access and infrastructure.

The discussion on the CBD, which seeks to give economic value to nature-derived DSI, will have a great impact on open science. It should be noted that national data sovereignty may deter global innovation by limiting access to vital data. Data ethics should be developed from a global perspective, leaving no country or community behind, and multiple fora working on different issues such as pathogens, crops, endangered species, and the open sea need to communicate with each other to achieve streamlined guidelines for benefit sharing from DSI.

FUNDING INFORMATION

This work is supported with JST-CREST (JPMJCR20H1) and AMED eAsia JRP (24jm0210113h0001).

COMPETING INTERESTS

The author is Head of DNA Data Bank of Japan (DDBJ), a part of INSDC. This work reflects only the author's view and INSDC is not responsible for the contents.

AUTHOR AFFILIATIONS

Masanori Arita  orcid.org/0000-0001-6706-0487

National Institute of Genetics, RIKEN Center for Sustainable Resource Science, JP

- Act on the Protection of Personal Information (Act No. 57 of, 2003): <https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en>.
- Alder, S.** (2024) 'HIPAA Explained', The HIPAA Journal Available at: <https://www.hipaajournal.com/hipaa-explained/>.
- Arthur, C.** (2023) 'Tech giants may be huge, but nothing matches big data,' *The Guardian*, 23 August 2013. Available at: <https://www.theguardian.com/technology/2013/aug/23/tech-giants-data>.
- Atokar** (2014) 'Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset', 15 September. Available at: <https://agkn.wordpress.com/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/>.
- BEIS** 2022. *Intelligent open science: a case study of viral genomic data sharing during the COVID-19 pandemic* (BEIS 2022/026). Available at: <https://www.gov.uk/government/publications/intelligent-open-science-viral-genomic-data-sharing-during-the-covid-19-pandemic>.
- CBD** *Conference of the parties (COP)*. Available at: <https://www.cbd.int/cop>.
- GDPR** (2016) *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC* (Document 32016R0679). EUR-Lex.
- Karsch-Mizrachi, I., et al.** (2024) 'The international nucleotide sequence database collaboration (INSDC): Enhancing global participation', *Nucleic Acids Res*, 53(D1), pp. D62–D66.
- Sheehan, N., Botta, F. and Leonelli, S.** (2024) 'Unrestricted versus regulated open data governance: A bibliometric comparison of sars-cov-2 nucleotide sequence databases', *Data Science J*, 23:29, pp. 1–29 Available at: <https://doi.org/10.5334/dsj-2024-029>
- UNESCO** *Open science*. Available at: <https://www.unesco.org/en/open-science>.
- WIPO** *Computer software*. Available at: <https://www.wipo.int/en/web/copyright/activities/software>.

TO CITE THIS ARTICLE:

Arita, M 2025 Data Sovereignty and Open Sharing: Reconciling Benefit-Sharing and Governance of Digital Sequence Information. *Data Science Journal*, 24: 9, pp. 1–5. DOI: <https://doi.org/10.5334/dsj-2025-009>

Submitted: 07 January 2025

Accepted: 14 February 2025

Published: 21 February 2025

COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.